TA ed.

BERICHTE
DER
ARBEITSGRUPPE TECHNOMATHEMATIK

FORSCHUNG - AUSBILDUNG - WEITERBILDUNG

BERICHT NR. 6

REGRESSION FÜR ELLIPSEN IN ACHSENPARALLELER LAGE

W. KRÜGER
200*

Interner Bericht

UNIVERSITÄT KAISERSLAUTERN
FACHBEREICH MATHEMATIK
ERWIN-SCHRÖDINGER-STRASSE
6750 KAISERSLAUTERN

OKTOBER 1984

REGRESSION FÜR ELLIPSEN

IN ACHSENPARALLELER LAGE

Dr. W. Krüger
Fachbereich Mathematik
Arbeitsgruppe Technomathematik
Universität Kaiserslautern

6750 Kaiserslautern

Die Aufgabe dieses Projektes*) war die Erstellung eines Verfahrens zur Bestimmung einer achsenparallelen Ausgleichsellipse. Dies bedeutet:

Gegeben ist eine Menge von Punkten $P_1 = (x_1, y_1), \ldots, P_n = (x_n, y_n)$, $n \geq 5$, des $\mathbb{R}^2$, die annähernd der Ellipsengleichung

$$(*) \qquad E : \frac{(x-x_M)^2}{a^2} + \frac{(y-y_M)^2}{b^2} = 1$$

genügen. Gesucht werden geeignete Werte für x_M, y_M, a, b.

Der nachfolgende Bericht gliedert sich in zwei Teile. Im ersten Teil wird die reine algorithmische Lösung des Problems (*) beschrieben, während im zweiten Teil die zugehörigen mathematischen Methoden hergeleitet werden. Der vorgeschlagene Lösungsalgorithmus ist ein Verfahren, welches sowohl mit den abgespeicherten Punkten als auch ohne Abspeicherung der Punkte arbeitet. Die Grundidee für den Algorithmus besteht darin, den Ausdruck

$$(**) \qquad D_1(a,b,x_M,y_M) = \sum_{i=1}^{n} \left(\frac{(x_i-x_M)^2}{a^2} + \frac{(y_i-y_M)^2}{b^2} - 1\right)^2$$

als Funktion in a, b, x_M, y_M zu minimieren und dadurch geeignete Werte für x_M, y_M, a, b zu bestimmen. Weshalb dieses Verfahren sinnvoll ist, also warum dadurch auch eine gute Lösung für die Minimierung von

$$(***) \qquad D_{Eu}^2(a,b,x_M,y_M) = \sum_{i=1}^{n} \mathrm{Min}_{(x,y)\in E}[(x-x_i)^2 + (y-y_i)^2]$$

bestimmt ist, wird im Teil II hergeleitet. Ebenso wird erläutert, warum der beschriebene Algorithmus überhaupt zur Lösung des Problems (*) führt.

---

*) Der technische Hintergrund dieses Projekts kann aus Wettbewerbsgründen nicht publiziert werden.

I. Algorithmus zur Lösung des Problems (*)

Gegeben seien

n : Anzahl der gemessenen Punkte $(n \geq 5)$

$(x_1, y_1), \ldots, (x_n, y_n)$: Koordinaten der gemessenen Punkte

1. Schritt: Berechne hieraus die Größen

$$S_{10} = \sum_{i=1}^{n} x_i \qquad S_{01} = \sum_{i=1}^{n} y_i$$

$$S_{20} = \sum_{i=1}^{n} x_i^2 \qquad S_{11} = \sum_{i=1}^{n} x_i y_i \qquad S_{02} = \sum_{i=1}^{n} y_i^2$$

$$S_{30} = \sum_{i=1}^{n} x_i^3 \qquad S_{21} = \sum_{i=1}^{n} x_i^2 y_i \qquad S_{12} = \sum_{i=1}^{n} x_i y_i^2 \qquad S_{03} = \sum_{i=1}^{n} y_i^3$$

$$S_{40} = \sum_{i=1}^{n} x_i^4 \qquad S_{22} = \sum_{i=1}^{n} x_i^2 y_i^2 \qquad S_{04} = \sum_{i=1}^{n} y_i^4$$

2. Schritt: Vereinbare folgende Funktionen

$$f_{10}(x) = \sum_{i=1}^{n} (x_i - x) = S_{10} - n \cdot x$$

$$f_{01}(y) = \sum_{i=1}^{n} (y_i - y) = S_{01} - n \cdot y$$

$$f_{20}(x) = \sum_{i=1}^{n} (x_i - x)^2 = S_{20} - 2 \cdot S_{10} \cdot x + n \cdot x^2$$

$$f_{11}(x,y) = \sum_{i=1}^{n} (x_i - x)(y_i - y) = S_{11} - S_{10} \cdot y - S_{01} \cdot x + n \cdot x \cdot y$$

$$f_{02}(y) = \sum_{i=1}^{n} (y_i - y)^2 = S_{02} - 2 \cdot S_{01} y + n \cdot y^2$$

$$f_{30}(x) = \sum_{i=1}^{n} (x_i - x)^3 = S_{30} - 3 \cdot S_{20} x + 3 \cdot S_{10} x^2 - n \cdot x^3$$

$$f_{21}(x,y) = \sum_{i=1}^{n} (x_i - x)^2 (y_i - y) = S_{21} - 2 \cdot S_{11} x + 2 \cdot S_{10} xy - S_{20} y + S_{01} x^2 - n \cdot x^2 \cdot y$$

$$f_{12}(x,y) = \sum_{i=1}^{n} (x_i - x)(y_i - y)^2 = S_{12} - 2 \cdot S_{11} y + 2 \cdot S_{01} xy - S_{02} x + S_{10} y^2 - n \cdot x \cdot y^2$$

$$f_{03}(y) = \sum_{i=1}^{n}(y_i-y)^3 = S_{03}-3\cdot S_{02}y+3\cdot S_{01}y^2-n\cdot y^3$$

$$f_{40}(x) = \sum_{i=1}^{n}(x_i-x)^4 = S_{40}-4\cdot S_{30}x+6\cdot S_{20}x^2-4\cdot S_{10}x^3+n\cdot x^4$$

$$f_{22}(x,y) = \sum_{i=1}^{n}(x_i-x)^2(y_i-y)^2 = S_{22}+S_{02}x^2+S_{20}y^2-2\cdot S_{12}x-2\cdot S_{21}y+4\cdot S_{11}$$
$$-2\cdot S_{10}\;xy^2-2\cdot S_{01}x^2y+n\cdot x^2$$

$$f_{04}(y) = \sum_{i=1}^{n}(y_i-y)^4 = S_{04}-4\cdot S_{03}y+6\cdot S_{02}y^2-4\cdot S_{01}y^3+n\cdot y^4$$

$$F_1(x,y) = \begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ f_{30}(x) & f_{12}(x,y) & f_{10}(x) \end{vmatrix}$$

$$F_2(x,y) = \begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ f_{21}(x,y) & f_{03}(y) & f_{01}(y) \end{vmatrix}$$

$$DXF_1(x,y) = 2\cdot\begin{vmatrix} 2f_{30}(x) & f_{12}(x,y) & f_{10}(x) \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ f_{30}(x) & 0 & 0 \end{vmatrix} - 2\cdot\begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{12}(x,y) & 0 & 0 \\ f_{30}(x) & f_{12}(x,y) & f_{10}(x) \end{vmatrix}$$

$$-\begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ 3f_{20}(x) & f_{02}(x,y) & n \end{vmatrix}$$

$$DXF_2(x,y) = -2\cdot\begin{vmatrix} 2f_{30}(x) & f_{12}(x,y) & f_{10}(x) \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ f_{21}(x,y) & f_{03}(y) & f_{01}(y) \end{vmatrix} - 2\cdot\begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{12}(x,y) & 0 & 0 \\ f_{21}(x,y) & f_{03}(y) & f_{01}(y) \end{vmatrix}$$

$$-2\cdot\begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ f_{11}(x,y) & 0 & 0 \end{vmatrix}$$

$$DYF_1(x,y) = -2 \cdot \begin{vmatrix} 0 & f_{21}(x,y) & 0 \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ f_{30}(x) & f_{12}(x,y) & f_{10}(x) \end{vmatrix} - 2 \cdot \begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{21}(x,y) & 2f_{03}(y) & f_{01}(y) \\ f_{30}(x) & f_{12}(x,y) & f_{10}(x) \end{vmatrix}$$

$$- 2 \cdot \begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ 0 & f_{11}(x,y) & 0 \end{vmatrix}$$

$$FYF_2(x,y) = -2 \cdot \begin{vmatrix} 0 & f_{21}(x,y) & 0 \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ f_{21}(x,y) & f_{03}(y) & f_{01}(y) \end{vmatrix} + 2 \cdot \begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{21}(x,y) & 2f_{03}(y) & f_{01}(y) \\ 0 & f_{03}(y) & 0 \end{vmatrix}$$

$$- \begin{vmatrix} f_{40}(x) & f_{22}(x,y) & f_{20}(x) \\ f_{22}(x,y) & f_{04}(y) & f_{02}(y) \\ f_{20}(x) & 3f_{02}(y) & n \end{vmatrix}$$

$$D(x,y) = DXF_1(x,y)\ DYF_2(x,y) - DYF_1(x,y)\ DXF_2(x,y)$$

$$g_1(x,y) = x - \frac{F_1(x,y)\ DYF_2(x,y) - F_2(x,y)\ DYF_1(x,y)}{D(x,y)}$$

$$g_2(x,y) = y - \frac{F_2(x,y)\ DXF_1(x,y) - F_1(x,y)\ DXF_2(x,y)}{D(x,y)}$$

3. Schritt: Bestimme einen Startwert x_o, y_o. Hierfür sind unter anderem folgende Alternativen möglich:

(i) $x_o = \frac{1}{n} S_{10} \qquad y_o = \frac{1}{n} S_{01}$

(ii) $x_o = x_{Min} + \frac{1}{2}(x_{Max} - x_{Min}) \qquad y_o = y_{Min} + \frac{1}{2}(y_{Max} - y_{Min})$,

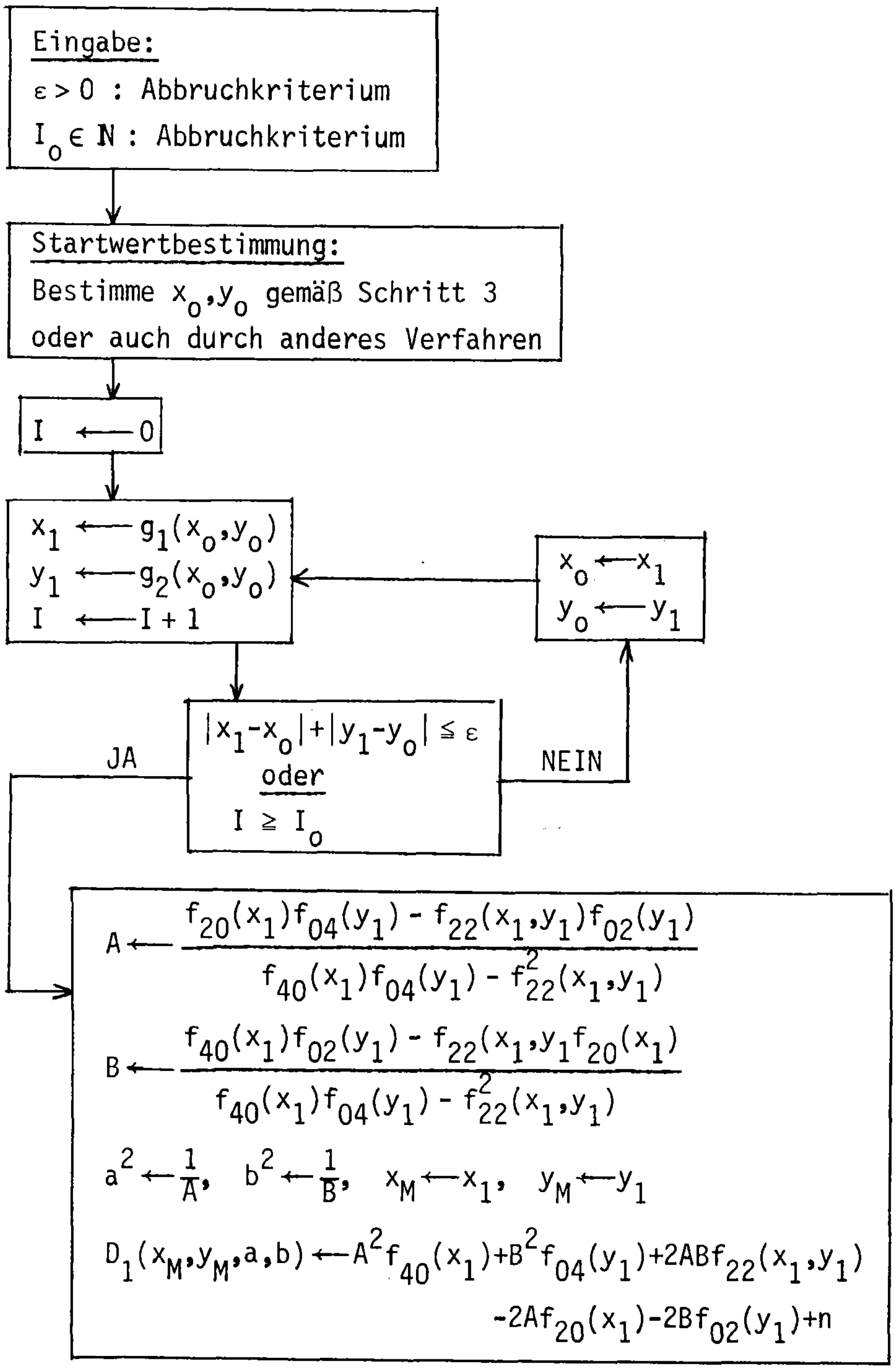
wobei

$x_{Max} = \text{Max}\ \{x_i; i=1,...,n\} \qquad x_{Min} = \text{Min}\ \{x_i; i=1,...,n\}$

$y_{Max} = \text{Max}\ \{y_i; i=1,...,n\} \qquad y_{Min} = \text{Min}\ \{y_i; i=1,...,n\}$

(iii) $x_o = \hat{x}_M$, $y_o = \hat{y}_M$, falls bereits eine Ausgleichsellipse berechnet wurde und eine neue Ellipse zu berechnen ist, deren Mittelpunkt sich nur wenig von dieser unterscheidet.

Nach diesen Vorbereitungen läßt sich der Algorithmus leicht durch folgenden Ablaufplan beschreiben:



Bemerkung:

1) Ich habe bei meiner Simulation stets $\varepsilon = 10^{-10}$ und $I_o = 7$ gewählt. Falls der Startwert (x_o, y_o) recht gut lag, reichte sogar $I_o = 4$.

2) Falls A oder B negativ sind oder falls $D_2(x_M, y_M, a, b)$ zu groß ist, muß man entweder mit einem besseren Startwert (x_o, y_o) neu iterieren oder die gemessenen Punkte als zu schlecht verwerfen.

3) Falls während der Rechnung die im zweiten Schritt definierte Größe $D(x,y)$ "numerisch" (d.h. im Sinne der Rechengenauigkeit) gleich Null ist, so kann man das Iterationsverfahren unmittelbar beenden und gleich zum letzten Schritt übergehen. (Dies trat bei meinen Simulationen allerdings nicht auf, daher habe ich dies auch nicht explizit in die Abbruchbedingung mit aufgenommen.)

II. Herleitung der mathematischen Methoden zur Lösung des Problems

1. Vorbetrachtungen

Bei gewissen Eichmessungen entsteht folgendes Problem:

Gegeben ist eine Menge von Punkten $P_1 = (x_1, y_1), \ldots, P_n = (x_n, y_n)$, $n \geq 5$, des $\mathbb{R}^2$, die annähernd der Ellipsengleichung

$$E : \frac{(x-x_M)^2}{a^2} + \frac{(y-y_M)^2}{b^2} = 1 \tag{1.1}$$

genügen. Gesucht werden geeignete Werte für x_M, y_M, a und b.

Bei einer Problemstellung in dieser Form ist es zunächst wichtig, über ein geeignetes Modell nachzudenken, in dessen Rahmen sich das obige Problem beschreiben läßt. Eine Möglichkeit zur Erklärung des obigen Problems im Rahmen eines Regressionsmodells besteht in folgender Modellannahme:

Zu vorgegebenen Winkeln $\varphi \in [0, 2\pi)$ beobachtet man Realisierungen $(x(\varphi), y(\varphi))$ einer Zufallsvariablen $(X(\varphi), Y(\varphi))$ mit

$$(X(\varphi), Y(\varphi)) = (x_M + a\cos\varphi, y_M + b\sin\varphi) + (\xi(\varphi), \eta(\varphi)), \tag{1.2}$$

wobei $(\xi(\varphi), \eta(\varphi))$ ein zweidimensionaler normalverteilter Zufallsvektor ist mit der Dichte

$$f_{(\varphi)}(x,y) = \frac{1}{2\pi\sigma\tau} \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \frac{y^2}{\tau^2}\right)\right) .$$

Nehmen wir nun an, daß alle Zufallsvariablen $(X(\varphi), Y(\varphi))$ für $\varphi \in [0, 2\pi)$ voneinander unabhängig sind und daß die Punkte $P_1, \ldots, P_n$ unabhängige Realisierungen von $(X(\varphi_1), Y(\varphi_1)), \ldots, (X(\varphi_n), Y(\varphi_n))$ zu vorgegebenen Winkeln $\varphi_1, \ldots, \varphi_n$ sind, dann liefert das Maximum-Likelihood-Verfahren eine Schätzung für x_M, y_M, a, b. Diese Schätzung läßt sich auch als Ergebnis einer Regressionsrechnung bei Parametrisierung über den Winkelbereich $[0, 2\pi)$ interpretieren. Die gemeinsame Dichte von $(X(\varphi_1), Y(\varphi_1)), \ldots, (X(\varphi_n), Y(\varphi_n))$ lautet

$$f_{(\varphi_1,\ldots,\varphi_n)}((x_1,y_1),\ldots,(x_n,y_n))$$
$$= \left(\frac{1}{2\pi\sigma\tau}\right)^n \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left\{\frac{1}{\sigma^2}(x_i-x_M-a\cos\varphi_i)^2+\frac{1}{\tau^2}(y_i-y_M-b\sin\varphi_i)^2\right\}\right).$$

Hieraus ergibt sich die Maximum-Likelihood-Funktion $L(x_M,y_M,a,b)$ für die Schätzung von x_M,y_M,a,b zu

$$L(x_M,y_M,a,b) = n\log 2\pi + n\log\sigma\tau + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-x_M-a\cos\varphi_i)^2 + \frac{1}{2\tau^2}\sum_{i=1}^{n}(y_i-y_M-b\sin\varphi_i)^2.$$

Die Bestimmung der Schätzwerte $\hat{x}_M,\hat{y}_M,\hat{a},\hat{b}$ ergibt sich dann aus der Lösung des linearen Gleichungssystems

$$\frac{\partial L}{\partial x_M}(\hat{x}_M,\hat{y}_M,\hat{a},\hat{b}) = \frac{\partial L}{\partial y_M}(\hat{x}_M,\hat{y}_M,\hat{a},\hat{b}) = \frac{\partial L}{\partial a}(\hat{x}_M,\hat{y}_M,\hat{a},\hat{b}) = \frac{\partial L}{\partial b}(\hat{x}_M,\hat{y}_M,\hat{a},\hat{b}) = 0.$$

Mit den Bezeichnungen

$$\hat{x} = \frac{1}{n}\sum_{i=1}^{n}x_i,\quad \hat{y} = \frac{1}{n}\sum_{i=1}^{n}y_i,\quad \hat{c}_1 = \frac{1}{n}\sum_{i=1}^{n}\cos\varphi_i,\quad \hat{s}_1 = \frac{1}{n}\sum_{i=1}^{n}\sin\varphi_i,$$

$$\hat{x}_c = \frac{1}{n}\sum_{i=1}^{n}x_i\cos\varphi_i,\quad \hat{y}_s = \frac{1}{n}\sum_{i=1}^{n}y_i\sin\varphi_i,\quad \hat{c}_2 = \frac{1}{n}\sum_{i=1}^{n}\cos^2\varphi_i,$$
$$\hat{s}_2 = \frac{1}{n}\sum_{i=1}^{n}\sin^2\varphi_i$$

ergibt sich folgende Lösung:

$$\hat{x}_M = \frac{\hat{x}\hat{c}_2-\hat{c}_1\hat{x}_c}{\hat{c}_2-\hat{c}_1^2}\qquad \hat{y}_M = \frac{\hat{y}\hat{s}_2-\hat{s}_1\hat{y}_s}{\hat{s}_2-\hat{s}_1^2}$$

$$\hat{a}^2 = \left(\frac{\hat{x}_c-\hat{c}_1\hat{x}}{\hat{c}_2-\hat{c}_1^2}\right)^2\qquad \hat{b}^2 = \left(\frac{\hat{y}_s-\hat{s}_1\hat{y}}{\hat{s}_2-\hat{s}_1^2}\right)^2\quad \text{für } (\hat{s}_2-\hat{s}_1^2)(\hat{c}_2-\hat{c}_1^2)\neq 0.$$

Der Vorteil bei diesem Verfahren besteht darin, daß man die Lösung sofort hinschreiben kann, dazu wird allerdings vorausgesetzt, daß die Winkel $\varphi_1,\ldots,\varphi_n$, unter denen die Punkte $P_1,\ldots,P_n$ gemessen werden, bekannt sind. Doch dies ist i.a. meist nicht der Fall; daher muß man sich für solche Fälle ein anderes Verfahren überlegen. Dabei stößt man zunächst auf prinzipielle Schwierigkeiten, die mit der Berechnung des Abstandes eines Punktes von einer Ellipse zusammenhängen. Ich will dies zunächst ausführlich für den üblichen ℓ^2-Abstand erläutern. Der Einfachheit halber betrachten wir dazu eine Ellipse E_o

$$E_o : \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

in Ursprungslage. Die Berechnung des Abstandes $D_{Eu}(x_1,y_1,a,b)$

$$D_{Eu}(x_1,y_1,a,b) := \mathrm{Min}_{(x,y) \in E_o} \left((x-x_1)^2 + (y-y_1)^2\right)^{1/2}$$

eines Punktes $P_1 = (x_1,y_1)$ von der Ellipse E_o in der euklidischen Metrik führt zur Nullstellenbestimmung eines Polynoms 4-ten Grades, d.h. $D_{Eu}(x_1,y_1,a,b)$ läßt sich nicht explizit angeben. Dies sieht man folgendermaßen:

Da die Ellipse E_o, vereinigt mit der von ihr umschlossenen Fläche, konvex ist, gibt es einen eindeutig bestimmten Punkt $P_o \in E_o$ so, daß

$$D_{Eu}(x_1,y_1,a,b) = (x_o-x_1)^2 + (y_o-y_1)^2)^{1/2}.$$

Für die Berechnung des Punktes P_o gelten folgende zwei Bedingungen:

(1.4) P_o muß auf der Ellipse E_o liegen.

(1.5) Die Gerade G durch P_o und P_1 muß senkrecht auf der Tangente T im Punkt P_o an die Ellipse E_o stehen.

Man erhält also folgende Gleichungen:

$$G : y = \frac{y_o-y_1}{x_o-x_1}x + \frac{y_1x_o-x_1y_1}{x_o-x_1} \qquad (P_1 \notin E_o,\ x_1 \neq 0)$$

$$T : y = -\frac{b^2}{a^2}\frac{x_o}{y_o}x + \frac{b^2}{y_o} \qquad \text{falls } y_o \neq 0$$

Falls $P_1 \in E_o$ oder $x_1 = 0$ oder $y_o = 0$, so ist die Bestimmung von P_o klar.

Also bedeuten die Bedingungen (1.4) und (1.5), daß

$$\frac{x_o^1}{a^2} + \frac{y_o^2}{b^2} = 1 \qquad \text{und} \qquad \frac{y_o - y_1}{x_o - x_1} \left(-\frac{b^2 x_o}{a^2 y_o}\right) = -1$$

d.h.

$$b^2 x_o^2 + a^2 y_o^2 = a^2 b^2 \qquad \text{und} \qquad (y_o - y_1) b^2 x_o = (x_o - x_1) a^2 y_o .$$

Untersuchen wir zunächst die zweite Bedingung:

$$(b^2 - a^2) x_o y_o + a^2 x_1 y_o - b^2 y_1 x_o = 0.$$

Die Invarianten dieser Kurve (in x_o und y_o) zweiter Ordnung (Kegelschnitt) sind:

$$\Delta = \begin{vmatrix} 0 & \frac{1}{2}(b^2-a^2) & -\frac{1}{2}b^2 y_1 \\ \frac{1}{2}(b^2-a^2) & 0 & \frac{1}{2}a^2 x_1 \\ -\frac{1}{2}b^2 y_1 & \frac{1}{2}a^2 x_1 & 0 \end{vmatrix} = \frac{1}{4}(b^2-a^2) a^2 b^2 x_1 y_1$$

$$\delta = \begin{vmatrix} 0 & \frac{1}{2}(b^2-a^2) \\ \frac{1}{2}(b^2-a^2) & 0 \end{vmatrix} = -\frac{1}{4}(b^2-a^2)^2$$

$$s = 0$$

Falls $x_1 \cdot y_1 \cdot (b^2 - a^2) \neq 0$, so liegt also eine Hyperbel vor, andernfalls ein Geradenpaar, welches auch zusammenfallen kann. Der Mittelpunkt dieser Hyperbel lautet

$$M = (x_M, y_M) \quad \text{mit} \quad x_M = \frac{a^2 x_1}{a^2 - b^2} \quad \text{und} \quad y_M = -\frac{b^2 y_1}{a^2 - b^2},$$

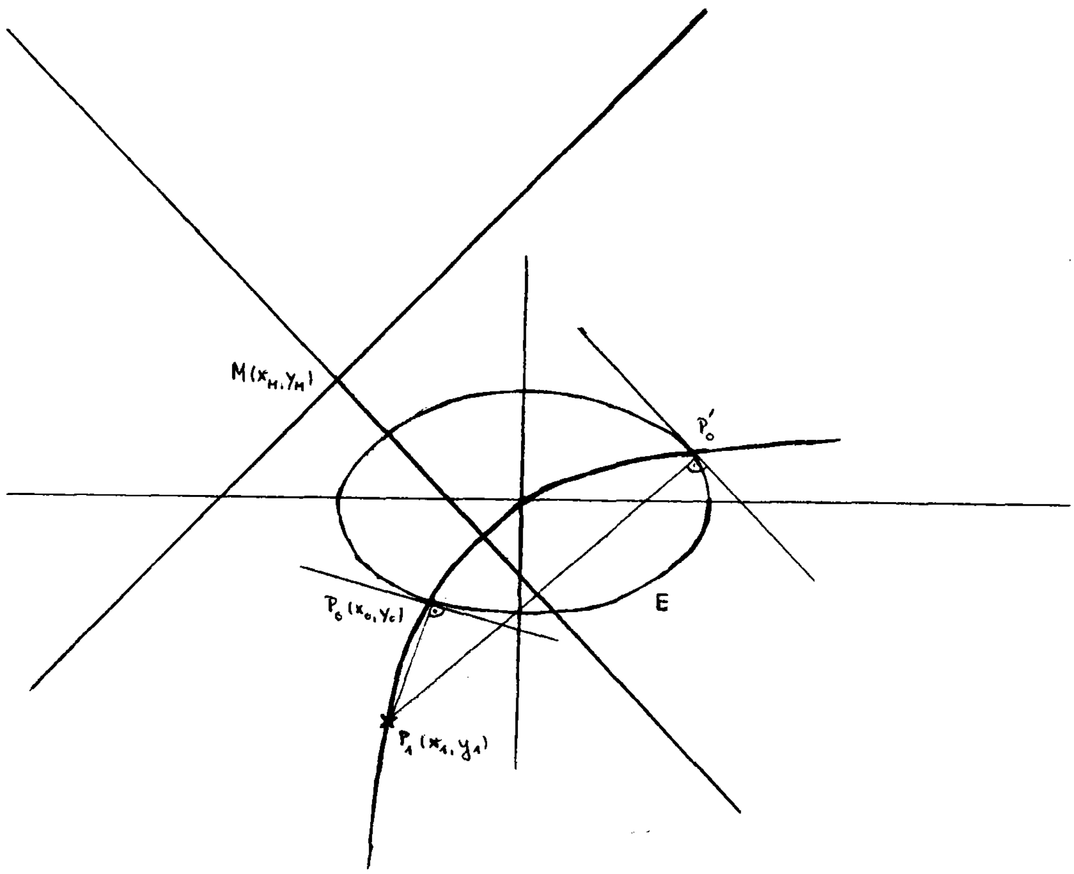
und der Drehkoeffizient ist gleich -1, d.h. Drehung um -45^o.

Die Normalform im entsprechend transformierten Koordinatensystem lautet dann

$$\frac{1}{2}(b^2-a^2)\bar{x}^2 - \frac{1}{2}(b^2-a^2)\bar{y}^2 - \frac{a^2b^2x_1y_1}{b^2-a^2} = 0,$$

also

$$\frac{\bar{x}^2}{\frac{2a^2b^2x_1y_1}{(b^2-a^2)^2}} - \frac{\bar{y}^2}{\frac{2a^2b^2x_1y_1}{(b^2-a^2)^2}} = 1$$



Diese Tatsache erklärt, daß die "Kopplung" der Gleichungen für (1.4) und (1.5) zu folgender Gleichung zur Bestimmung von x_o führt:

$$e^4x_o^4-2e^2a^2x_1x_o^3+x_o^2(a^4x_1^2+a^2b^2y_1^2-a^2e^2)+2a^4e^2x_1x_o-a^6x_1^2 = 0,$$

wobei $e^2 = a^2-b^2$.

Aber auch andere Normen vereinfachen das Problem nicht wesentlich. Man hat immer noch die Schwierigkeit, die Lage des Punktes relativ zur Ellipse kennen zu müssen.

Beispiel: Maximumsnorm

Der Abstand zweier Punkte $P_1 = (x_1, y_1)$ und $P_2 = (x_2, y_2)$ in der Maximumsnorm ist definiert als

$$D_{Max}(P_1, P_2) = \text{Max} \{|x_2 - x_1|, |y_2 - y_1|\}.$$

Berechnet man damit den Abstand eines Punktes $P_1 = (x_1, y_1)$ von der Ellipse E_o, so erhält man zur Bestimmung des "Fußpunktes" $P_o = (x_o, y_o)$ die Gleichungen

$$y_o - y_1 = x_o - x_1 \quad \text{und} \quad \frac{x_o^2}{a^2} + \frac{y_o^2}{b^2} = 1.$$

Seien x_{o1} und x_{o2} die Lösungen der quadratischen Gleichung

$$x_o^2(a^2+b^2) + 2x_o(y_1-x_1) + a^2(y_1-x_1)^2 - a^2b^2 = 0$$

und

$$y_{o11} = + \sqrt{b^2 - \frac{b^2}{a^2} x_{o1}^2} \qquad y_{o12} = - \sqrt{b^2 - \frac{b^2}{a^2} x_{o1}^2}$$

$$y_{o21} = + \sqrt{b^2 - \frac{b^2}{a^2} x_{o2}^2} \qquad y_{o22} = - \sqrt{b^2 - \frac{b^2}{a^2} x_{o2}^2},$$

dann gilt für den Fußpunkt P_o

$$P_o \in \{(x_{o1}, y_{o11}), (x_{o1}, y_{o12}), (x_{o2}, y_{o21}), (x_{o2}, y_{o22}), (0,b), (0,-b), (a,0), (-a,0)\}.$$

Man muß also zur Bestimmung von P_o die Lage von P_o relativ zur Ellipse E_o kennen.

Daher muß man sich zur Berechnung des Abstandes eines Punktes P_1 von der Ellipse E eine andere Methode einfallen lassen, als nur die Abänderung der entsprechenden Norm.

2. Konstruktion eines Abstandes über affine Abbildungen

Jede Ellipse läßt sich als affines Bild eines Kreises darstellen. Betrachten wir der Einfachheit halber wieder die Ellipse E_o in Ursprungslage, d.h.

$$E_o = \{(x,y) \in \mathbb{R}^2 , \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\}$$

und die affine Abbildung

$$V : \mathbb{R}^2 \to \mathbb{R}^2 \text{ mit } V(x,y) = (x, \frac{b}{a} y),$$

dann gilt bekanntlich $V(K) = E_o$, wobei

$$K = \{(x,y) \in \mathbb{R}^2; x^2+y^2 = a^2\}.$$

Berechnet man nun den Abstand eines Punktes $P_1 = (x_1,y_1)$ von der Ellipse E_o durch

$$D_{Aff} = \operatorname{Min}_{(x,y) \in V^{-1}(E_o)} ((x-\bar{x}_1)^2 + (y-\bar{y}_1))^{1/2},$$

wobei $(\bar{x}_1,\bar{y}_1) = V^{-1}((x_1,y_1))$, so erhält man

$$D^2_{Aff} = ((x_1^2 + \frac{a^2}{b^2} y_1^2)^{1/2} - a)^2$$

$$= a^2((\frac{x_1^2}{a^2} + \frac{y_1^2}{b^2})^{1/2} - 1)^2.$$

Für eine Ellipse E in allgemeiner Lage bedeutet dies also

$$D^2_{Aff} (x_M,y_M,a,b) = a^2((\frac{(x_1-x_M)^2}{a^2} + \frac{(y_1-y_M)^2}{b^2})^{1/2} - 1)^2$$

$$\leq a^2 \left| \frac{(x_1-x_M)^2}{a^2} + \frac{(y_1-y_M)^2}{b^2} - 1 \right|$$

3. Konstruktion eines Abstandes über achsenähnliche Ellipsen

Wir gehen diesmal gleich von einer Ellipse E in allgemeiner Lage (vgl. 1.1) aus. Legt man durch den Punkt $P_1 = (x_1, y_1)$ eine Ellipse E_1 mit demselben Mittelpunkt wie E und den Achsen A,B so, daß

$$A \cdot b = B \cdot a,$$

so führt dies für $(x_1, y_1) \neq (x_M, y_M)$ wegen

$$\frac{(x_1 - x_M)^2}{A^2} + \frac{(y_1 - y_M)^2}{B^2} = 1$$

zu

$$A^2 = (x_1 - x_M)^2 + \frac{a^2}{b^2}(y_1 - y_M)^2$$

und

$$B^2 = \frac{b^2}{a^2}(x_1 - x_M)^2 + (y_1 - y_M)^2,$$

also

$$A^2 = a^2\left(\frac{(x_1 - x_M)^2}{a^2} + \frac{(y_1 - y_M)^2}{b^2}\right) \quad \text{und} \quad B^2 = b^2\left(\frac{(x_1 - x_M)^2}{a^2} + \frac{(y_1 - y_M)^2}{b^2}\right).$$

Also ist der oben über affine Abbildungen konstruierte Abstand nichts anderes als

$$D^2_{Aff}(x_m, y_m, a, b) = (A-a)^2.$$

Man sieht hier eine gewisse Unsymmetrie in a und b. Dies ist jedoch nur auf den ersten Blick der Fall. Da $A = B \cdot \frac{a}{b}$, gilt nämlich

$$(A-a)^2 = \frac{a}{b}(A-a)(B-b).$$

Dennoch bleibt das Problem, daß in diesem Ausdruck eine Wurzel steht. Diese Problematik für die Lösung von (1.1) unter gewissen Randbedingungen werden wir im nächsten Abschnitt ausführlich behandeln. Wir wollen nun noch zwei weitere Abstände einführen und diese mit dem üblichen euklidischen Abstand vergleichen.

Dazu seien

$$D_1(x_M,y_M,a,b) := \left(\frac{A^2-a^2+B^2-b^2}{a^2+b^2}\right)^2$$

und

$$D_2(x_M,y_M,a,b) := (A^2-a^2)(B^2-b^2).$$

Wegen der obigen Beziehung für A^2 und B^2 gilt dann

$$D_1(x_M,y_M,a,b) = \frac{D_2(x_M,y_M,a,b)}{a^2b^2}$$

und

$$D_2(x_M,y_M,a,b) = a^2b^2\left(\frac{(x_1-x_M)^2}{a^2} + \frac{(y_1-y_M)^2}{b^2} - 1\right)^2.$$

<u>Lemma 1:</u>

Sei $D_{Eu}(x_M,y_M,a,b)$ der übliche ℓ^2-Abstand eines Punktes $P_1 = (x_1,y_1)$ von der Ellipse E, d.h.

$$D_{Eu}(x_M,y_M,a,b) = \mathrm{Min}_{(x,y)\in E}((x-x_1)^2 + (y-y_1)^2)^{1/2},$$

dann gelten folgende Abschätzungen:

$$D_{Eu}^4(x_M,y_M,a,b) \leq \frac{a^2}{b^2} D_2(x_M,y_M,a,b)$$

und

$$D_2(x_M,y_M,a,b) \leq \frac{a^2}{b^2} (D_{Eu}(x_M,y_M,a,b) + 2b)^2 \cdot D_{Eu}^2(x_M,y_M,a,b).$$

<u>Beweis:</u>

Zunächst gilt

$$|B-b| \leq D_{Eu}(x_M,y_M,a,b) \leq |A-a|,$$

also folgt

$$D_{Eu}^4(x_M,y_M,a,b) \leq (A-a)^4 = \frac{a^2}{b^2}(A-a)^2(B-b)^2$$

$$\leq \frac{a^2}{b^2}(A^2-a^2)(B^2-b^2), \qquad \text{da } \mathrm{sgn}(A-a) = \mathrm{sgn}\,(B-b).$$

Umgekehrt ist wegen

$$\begin{aligned} D_2(x_M,y_M,a,b) &= (A^2-a^2)(B^2-b^2) \\ &= \frac{a^2}{b^2}(B-b)^2(B+b)^2 \\ &\leq \frac{a^2}{b^2}(B-b)^2(|B-b|+2b)^2 \\ &\leq \frac{a^2}{b^2}\, D_{Eu}^2(x_M,y_M,a,b)(D_{Eu}(x_M,y_M,a,b)+2b)^2 \qquad \square \end{aligned}$$

Aus Lemma 1 folgt nun, daß aus Kenntnis des Gesamtabstandes D_2,

$$D_2 = \sum_{i=1}^{n} D_2(x_M,y_M,a,b,P_i),$$

aller Punkte $P_1,\dots,P_n$ im Sinne des D_2-Abstandes der Abstand D_{Eu},

$$D_{Eu} = \sum_{i=1}^{n} D_{Eu}(x_M,y_M,a,b,P_i),$$

aller Punkte von der Ellipse abgeschätzt werden kann. Es gilt nämlich

$$(3.1) \qquad D_{Eu} \leq \frac{a^2}{b^2}\, n^{3/4} D_2^{1/4}$$

und

$$(3.2) \qquad D_2 \leq \frac{a^2}{b^2}(D_{Eu}(D_{Eu}+2b))^2,$$

denn wegen Lemma 1 und der Hölderschen Ungleichung ist

$$\begin{aligned} D_{Eu} &= \sum_{i=1}^{n} D_{Eu}(x_M,y_M,a,b,P_i) \\ &\leq (\sum_{i=1}^{n} 1)^{3/4} \cdot (\sum_{i=1}^{n} D_{Eu}(x_M,y_M,a,b,P_i)^4)^{1/4} \\ &\leq n^{3/4} \cdot \frac{a^2}{b^2}\, D_2^{1/4} \end{aligned}$$

und umgekehrt gilt

$$
\begin{aligned}
D_2 &\leq \frac{a^2}{b^2}\sum_{i=1}^{n}(D_{Eu}(x_M,y_M,a,b,P_i)+2b)^2 D_{Eu}^2(x_M,y_M,a,b,P_i)\\
&\leq \frac{a^2}{b^2}\Big(\big(\sum_{i=1}^{n} D_{Eu}(x_M,y_M,a,b,P_i)\big)^4 + 2b\big(\sum_{i=1}^{n} D_{Eu}(x_M,y_M,a,b,P_i)\big)^3\\
&\qquad + \big(\sum_{i=1}^{n} D_{Eu}(x_M,y_M,a,b,P_i)\big)^2\Big)\\
&= \frac{a^2}{b^2}(D_{Eu}(D_{Eu}+2b))^2 .
\end{aligned}
$$

Ähnlich leicht läßt sich auch der Gesamtabstand aller Punkte $P_1,\ldots,P_n$ im Sinne des D_{Aff}-Abstandes mit dem D_2-Abstand vergleichen. Es gilt nämlich

$$D_{Aff}^2(x_M,y_M,a,b) \leq a^4 D_2(x_M,y_M,a,b) \tag{3.3}$$

und

$$D_2(x_M,y_M,a,b) \leq D_{Aff}^2(x_M,y_M,a,b)\left(\frac{b}{a}D_{Aff}(x_M,y_M,a,b)+b\right)^2 \tag{3.4}$$

und hieraus ergeben sich für die Gesamtabstände D_{Aff} und D_2 ähnlich wie in (3.1) und (3.2) die Beziehungen

$$D_{Aff} \leq a^2 \cdot n^{1/2}\, D_2^{1/2} \tag{3.5}$$

und

$$D_2 \leq D_{Aff}^2\left(\frac{b}{a}D_{Aff}+b\right)^2 . \tag{3.6}$$

Stellen wir die in (3.1), (3.2), (3.5) und (3.6) gewonnenen Resultate zusammen, so erhalten wir also

<u>Lemma 2:</u>

Gegeben seien die Punkte $P_1 = (x_1,y_1),\ldots,P_n = (x_n,y_n)$ und eine Ellipse $E = \{(x,y) \in \mathbb{R}^2 ; \frac{(x-x_M)^2}{a^2} + \frac{(y-y_M)^2}{b^2} = 1\}$, dann gilt für die Gesamtabstände aller Punkte $P_1,\ldots,P_n$ von der Ellipse E im Sinne des D_2-, D_{Aff}- und D_{Eu}-Abstandes folgende Beziehung:

$$D_{Eu} \leq \frac{a^2}{b^2} n^{3/4} D_2^{1/4} \quad , \quad D_2 \leq \frac{a^2}{b^2} (D_{Eu}(D_{Eu}+2b))^2$$

und

$$D_{Aff} \leq a^2 n^{1/2} D_2^{1/2} \quad , \quad D_2 \leq (D_{Aff}(\frac{b}{a} D_{Aff}+b))^2 .$$

Beim Vergleich zwischen D_{Eu} und D_2 stört zunächst etwas die 4-te Potenz von D_{Eu}. Dies ist aber kein eigentliches Problem, denn die zur Maximum Likelihood-Schätzung gehörige Metrik ist das Quadrat des ℓ^2-Abstandes. Damit gelten für $D_{ML} = D_{Eu}^2$ die Abschätzungen

$$D_{ML}^2(x_M,y_M,a,b) \leq \frac{a^2}{b^2} D_2(x_M,y_M,a,b)$$

und somit

$$\begin{aligned} D_{ML} &= \sum_{i=1}^{n} D_{Eu}^2(x_M,y_M,a,b,P_i) \\ &\leq (\sum_{i=1}^{n} 1)^{1/2} \cdot (\sum_{i=1}^{n} D_{Eu}^4(x_M,y_M,a,b,P_i))^{1/2} \\ &\leq n^{1/2} \cdot \frac{a^2}{b^2} D_2^{1/2} . \end{aligned}$$

Wegen der Beziehung

$$D_2(x_M,y_M,a,b) = a^2b^2 \, D_1(x_M,y_M,a,b) ,$$

folgt aus Lemma 2 auch leicht eine Abschätzung für die Gesamtabstände aller Punkte $P_1,\dots,P_n$ von der Ellipse E im Sinne des D_{Aff}- und D_{ML}-Abstandes zum D_1-Abstand:

$$D_{ML} \leq a^2 \cdot \frac{a}{b} \quad (n \cdot D_1)^{1/2}$$

(3.7)

$$D_{Aff} \leq a^3 b \cdot (n \cdot D_1)^{1/2} .$$

Also gilt insbesondere für die Summe der mittleren ℓ^2-Abstandsquadrate

(3.8) $$D_{ML} = \frac{1}{n} \sum_{i=1}^{n} D_{En}^2(x_M,y_M,a,b,P_i) \leq \frac{1}{\sqrt{n}} a^2 \frac{a}{b} D_1^{1/2}$$

Diese Größe kann später ein Maß für die Güte der berechneten Ellipse und damit der gemessenen Punkte sein.

Nach diesen Vorüberlegungen müssen wir uns nun für einen der besprochenen Abstände entscheiden. Da eine Randbedingung für die Lösung von (1.1) darin bestand, ein Verfahren zu entwickeln, welches ohne Abspeicherung der Punkte $P_1,\dots,P_n$ arbeitet, sondern nur (wenige) Größen verwendet, welche "on line" berechnet werden können, scheidet der Abstand D_{Aff} aus. Er würde zudem numerisch schwieriger zu behandeln sein als die Abstände D_1 und D_2. Der Unterschied zwischen den verbleibenden Abständen D_1 und D_2 ist nur eine von a und b abhängige Skalierung, welche allerdings die numerische Behandlung von D_2 gegenüber D_1 wesentlich erschwert, denn die Größen a und b treten bei D_2 in höheren Potenzen als bei D_1 auf. Daher ist die Verwendung des D_1-Abstandes zur Lösung des Problems (1.1) am sinnvollsten. Bevor wir auf die Lösung des Problems (1.1) mit Hilfe dieses Abstandes eingehen, wollen wir uns noch an Hand von Äquiniveaulinien den Skalierungsunterschied zwischen D_1 und D_2 klarmachen. Dazu berechnen wir zum Abstand d^2 ($d^2 < 1$) jeweils die Äquiniveaulinien zum Abstand D_1 und D_2 und bilden das Verhältnis der Abstände von zwei zugehörigen Punkten im Inneren und Äußeren der Ellipse auf diesen Linien im Sinne des Euklidischen Abstandes. Es ergibt sich

$$\frac{D_2A}{D_2I} = \frac{\sqrt{1+\frac{d}{ab}}-1}{1-\sqrt{1-\frac{d}{ab}}} \geq 1$$

und

$$\frac{D_1A}{D_1I} = \frac{\sqrt{1+d}-1}{1-\sqrt{1-d}} \geq 1 .$$

Dies heißt, daß sowohl für den D_2- als auch für den D_1-Abstand die später konstruierte Ellipse die Tendenz haben wird, daß mehr Punkte im Äußeren als im Inneren der Ellipse liegen. Für den D_2-Abstand wird dieser Effekt nicht ganz so stark sein wie für den D_1-Abstand. Falls d sehr klein ist, d.h. die gemessenen Punkte nicht allzu sehr verrauscht sind, kann dieser Effekt sowohl für den D_1-Abstand für den D_2-Abstand vernachlässigt werden.

4. Bestimmung einer "Ausgleichsellipse" im Sinne des D_1-Abstandes

Wie in (1.1) seien wieder n Punkte $P_1 = (x_1,y_1),\ldots,P_n(x_n,y_n)$ gegeben. Wir lösen nun das folgende Problem:

Bestimme x_M, y_M, a^2, b^2 so, daß

$$D_1(x_M,y_M,a,b) = \sum_{i=1}^{n} \left(\frac{(x_i-x_M)^2}{a^2} + \frac{(y_i-y_M)^2}{b^2}\right)^2 \tag{4.1}$$

minimal wird.

Dazu verwenden wir folgende Bezeichnungen:

$$S_{k\ell} = \sum_{i=1}^{n} x_i^k y_i^\ell, \text{ wobei } k,\ell \in \mathbb{N}_o \tag{4.2}$$

sowie die Funktionen

$$f_{k\ell}(x,y) = \sum_{i=1}^{n} (x_i-x)^k (y_i-y)^\ell, \text{ wobei } k,\ell \in \mathbb{N}_o\,. \tag{4.3}$$

Sei

$$M = \{(1,0,(0,1),(2,0),(1,1),(0,2),(3,0),(1,2),(2,1),(0,3),(4,0),(2,2),(0,4)\},$$

dann genügt es, $S_{k\ell}$ und $f_{k\ell}$ für $(k,\ell)\in M$ zu betrachten. Damit ergibt sich als notwendige Bedingung für die Bestimmung von x_M, y_M, a und b:

$$\frac{\partial D_1}{\partial a} = \frac{2}{a^3} \cdot 2 \cdot \left(\frac{f_{40}(x_M,y_M)}{a^2} + \frac{f_{22}(x_M,y_M)}{b^2} - f_{20}(x_M,y_M)\right) = 0$$

$$\frac{\partial D_1}{\partial b} = \frac{2}{b^3} \cdot 2 \cdot \left(\frac{f_{22}(x_M,y_M)}{a^2} + \frac{f_{04}(x_M,y_M)}{b^2} - f_{02}(x_M,y_M)\right) = 0$$

(4.4)

$$\frac{\partial D_1}{\partial x_M} = \frac{2}{a^2} \cdot 2 \cdot \left(\frac{f_{30}(x_M,y_M)}{a^2} + \frac{f_{12}(x_M,y_M)}{b^2} - f_{10}(x_M,y_M)\right) = 0$$

$$\frac{\partial D_1}{\partial y_M} = \frac{2}{b^2} \cdot 2 \cdot \left(\frac{f_{21}(x_M,y_M)}{a^2} + \frac{f_{03}(x_M,y_M)}{b^2} - f_{01}(x_M,y_M)\right) = 0$$

also mit $A = \frac{1}{a^2}$ und $B = \frac{1}{b^2}$ bedeutet dies

(4.5)

$$A\ f_{40}(x_M,y_M) + B\ f_{22}(x_M,y_M) - f_{20}(x_M,y_M) = 0$$

$$A\ f_{22}(x_M,y_M) + B\ f_{04}(x_M,y_M) - f_{02}(x_M,y_M) = 0$$

$$A\ f_{30}(x_M,y_M) + B\ f_{12}(x_M,y_M) - f_{10}(x_M,y_M) = 0$$

$$A\ f_{21}(x_M,y_M) + B\ f_{03}(x_M,y_M) - f_{01}(x_M,y_M) = 0.$$

Für unsere weiteren Überlegungen benötigen wir noch die Ableitungen der Funktionen $f_{k,\ell}$ für $(k,\ell) \in M$. Seien

$$f_{00}(x,y) = n$$

und

$$f_{k\ell}(x,y) = 0 \quad \text{für } k < 0 \text{ oder } \ell < 0,$$

dann gilt

$$\frac{\partial f_{k\ell}(x,y)}{\partial x} = -k \cdot f_{k-1,\ell}(x,y) \quad \text{und} \quad \frac{\partial f_{k,\ell}(x,y)}{\partial y} = -\ell \cdot f_{k,\ell-1}(x,y)$$

für alle $(k,\ell) \in M$.

Weiterhin benötigen wir für das später verwendete Iterationsverfahren ein zweidimensionales Newton-Verfahren, d.h. die Nullstellenbestimmung von

$$h_1(x,y) = h_2(x,y) = 0 \quad , \quad h_1,h_2 : \mathbb{R}^2 \to \mathbb{R} \text{ genügend glatt,}$$

erfolgt durch folgendes Iterationsverfahren:

$$g(x,y) = \binom{x}{y} - \frac{1}{H'(x,y)} H(x,y),$$

wobei

$$H(x,y) = \begin{pmatrix} h_1(x,y) \\ h_2(x,y) \end{pmatrix}$$

und

$$\frac{1}{H'(x,y)} = \begin{pmatrix} \frac{\partial h_1(x,y)}{\partial x} & \frac{\partial h_1(x,y)}{\partial y} \\ \frac{\partial h_2(x,y)}{\partial x} & \frac{\partial h_2(x,y)}{\partial y} \end{pmatrix}^{-1}$$

$$= \frac{1}{\frac{\partial h_1(x,y)}{\partial x} \frac{\partial h_2(x,y)}{\partial y} - \frac{\partial h_1(x,y)}{\partial y} \frac{\partial h_2(x,y)}{\partial x}} \cdot \begin{pmatrix} \frac{\partial h_2(x,y)}{\partial y} & -\frac{\partial h_1(x,y)}{\partial y} \\ -\frac{\partial h_2(x,y)}{\partial x} & \frac{\partial h_1(x,y)}{\partial x} \end{pmatrix},$$

also

$$g(x,y) = \binom{x}{y} - \frac{1}{\frac{\partial h_1(x,y)}{\partial x} \cdot \frac{\partial h_2(x,y)}{\partial y} - \frac{\partial h_1(x,y)}{\partial y} \cdot \frac{\partial h_2(x,y)}{\partial x}} \cdot \begin{pmatrix} h_1(x,y) \frac{\partial h_2(x,y)}{\partial y} - h_2(x,y) \frac{\partial h_1(x,y)}{\partial y} \\ h_2(x,y) \frac{\partial h_1(x,y)}{\partial x} - h_1(x,y) \frac{\partial h_2(x,y)}{\partial x} \end{pmatrix}$$

für $\frac{\partial h_1(x,y)}{\partial x} \frac{\partial h_2(x,y)}{\partial y} - \frac{\partial h_1(x,y)}{\partial y} \frac{\partial h_2(x,y)}{\partial x} \neq 0.$

Zusätzlich benötigen wir folgende einfache Beziehung.

Lemma 3

Gegeben seien folgende Gleichungssysteme in A und B

$$\text{(I)}\quad \begin{aligned} Aa_1 + Ba_2 &= a_3 \\ Ab_1 + Bb_2 &= b_3 \end{aligned} \qquad \text{und} \qquad \text{(II)}\quad \begin{aligned} Ac_1 + Bc_2 &= c_3 \\ Ad_1 + Bd_2 &= d_3. \end{aligned}$$

Dann gilt:

Falls $a_1b_2 - b_1a_2 \neq 0$, so ist das gesamte Gleichungssystem (I) und (II) in A und B lösbar genau dann, wenn

$$(*) \qquad \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ d_1 & d_2 & d_3 \end{vmatrix} = 0.$$

Insbesondere ist in diesem Fall das gesamte Gleichungssystem eindeutig lösbar.

Beweis:

Falls $a_1b_2-b_1a_2 \neq 0$, so ist das Gleichungssystem (I) eindeutig lösbar durch

$$A_1 = \frac{a_3b_2 - a_2b_3}{a_1b_2 - a_2b_1} \qquad \text{und} \qquad B_1 = \frac{a_1b_3 - a_3b_1}{a_1b_2 - a_2b_1},$$

und es reicht zu zeigen, daß A_1 und B_1 auch Lösungen von (II) sind genau dann, wenn (*) gilt. Nun sind aber A_1 und B_1 auch Lösungen von (II), genau dann, wenn

$$c_1 \cdot \begin{vmatrix} a_3 & a_2 \\ b_3 & b_2 \end{vmatrix} + c_2 \cdot \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} - c_3 \cdot \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = 0$$

und

$$d_1 \cdot \begin{vmatrix} a_3 & a_2 \\ b_3 & b_2 \end{vmatrix} + d_2 \cdot \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} - d_3 \cdot \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = 0.$$

Dies ist aber äquivalent zu (*), wie man unmittelbar durch Entwicklung der beiden Determinanten in (*) nach der letzten Zeile sieht. $\square$

Zur Anwendung dieses Lemmas auf das Gleichungssystem (4.5) müssen wir also überprüfen, ob

$$f_{40}(x_M,y_M) \cdot f_{04}(x_M,y_M) - f_{22}^2(x_M,y_M) \neq 0.$$

Wegen der Cauchy-Schwartz'schen Ungleichung gilt aber

$$f_{22}^2(x,y) \leq f_{40}(x,y) \cdot f_{04}(x,y) \qquad \text{für alle } x,y \in \mathbb{R}$$

und das Gleichheitszeichen gilt, falls nicht alle Punkte auf einer Parallelen zur x-Achse oder y-Achse liegen, genau dann, falls es ein $\alpha = \alpha(x,y) > 0$ gibt, so daß

$$(y_i - y)^2 = \alpha(x_i - x)^2 \qquad \text{für alle } i=1,\dots,n.$$

Dies bedeutet aber, daß

$$y_i - y = \pm\sqrt{\alpha}(x_i - x) \qquad \text{für alle } i=1,\dots,n,$$

also liegen alle Punkte $P_1,\dots,P_n$ auf einer der beiden Geraden

$$g_1(t) = \alpha\sqrt{t} + y - \sqrt{\alpha}x \quad \text{oder} \quad g_2(t) = -\sqrt{\alpha}t + y + \sqrt{\alpha}x \quad ,$$

deren gemeinsamer Schnittpunkt gerade der Punkt $P = (x,y)$ ist. Schließt man also aus, daß alle Punkte auf einer oder zwei Geraden liegen, so gilt stets

$$f_{40}(x,y) \cdot f_{04}(x,y) > f_{22}^2(x,y) \qquad \text{für alle } x,y \in \mathbb{R}$$

und daher ist wegen Lemma 3 das Gleichungssystem (4.5) äquivalent zu

$$F_1(x_M,y_M) = \begin{vmatrix} f_{40}(x_M,y_M) & f_{22}(x_M,y_M) & f_{20}(x_M,y_M) \\ f_{22}(x_M,y_M) & f_{04}(x_M,y_M) & f_{02}(x_M,y_M) \\ f_{30}(x_M,y_M) & f_{12}(x_M,y_M) & f_{10}(x_M,y_M) \end{vmatrix} = 0$$

$$(4.6) \qquad F_2(x_M,y_M) = \begin{vmatrix} f_{40}(x_M,y_M) & f_{22}(x_M,y_M) & f_{20}(x_M,y_M) \\ f_{22}(x_M,y_M) & f_{04}(x_M,y_M) & f_{02}(x_M,y_M) \\ f_{21}(x_M,y_M) & f_{03}(x_M,y_M) & f_{01}(x_M,y_M) \end{vmatrix} = 0$$

$$A = \frac{f_{20}(x_M,y_M)\ f_{04}(x_M,y_M) - f_{22}(x_M,y_M)\ f_{02}(x_M,y_M)}{f_{40}(x_M,y_M)\ f_{04}(x_M,y_M) - f_{22}^2(x_M,y_M)}$$

$$B = \frac{f_{40}(x_M,y_M)\ f_{02}(x_M,y_M) - f_{20}(x_M,y_M)\ f_{22}(x_M,y_M)}{f_{40}(x_M,y_M)\ f_{04}(x_M,y_M) - f_{22}^2(x_M,y_M)} \ .$$

Damit haben wir das nichtlineare Gleichungssystem (4.5) in x_M, y_M und A,B entkoppelt. Es bleibt zur Lösung von (1.1) die Bestimmung der gemeinsamen Nullstellen von F_1 und F_2. Dazu verwenden wir das oben beschriebene Newton-Verfahren. Somit ergibt sich folgende Iterationsvorschrift:

$$\begin{pmatrix} x_M^{(\nu+1)} \\ y_M^{(\nu+1)} \end{pmatrix} = g \begin{pmatrix} x_M^{(\nu)} \\ y_M^{(\nu)} \end{pmatrix} \qquad \nu=0,1,\ldots ,$$

wobei

$$g\binom{x}{y} = \binom{x}{y} - \frac{1}{\frac{\partial F_1(x,y)}{\partial x}\frac{\partial F_2(x,y)}{\partial y} - \frac{\partial F_1(x,y)}{\partial y}\frac{\partial F_2(x,y)}{\partial x}} \cdot \begin{pmatrix} F_1(x,y)\,\frac{\partial F_2(x,y)}{y} - F_2(x,y)\,\frac{\partial F_1(x,y)}{y} \\ F_2(x,y)\,\frac{\partial F_1(x,y)}{\partial x} - F_1(x,y)\,\frac{\partial F_2(x,y)}{\partial x} \end{pmatrix}$$

Für die Wahl eines Startwertes werden wir im nächsten Abschnitt mehrere Verfahren vorstellen. Nach einer entsprechenden Anzahl von Iterationen bestimmt man A und B aus den beiden letzten Gleichungen von (4.7). Dabei haben numerische Experimente gezeigt, daß man mit weniger als 6 Iterationen auskommt. Die einzige Voraussetzung dabei war, einen geeigneten Startwert vorzugeben.

5. A-priori Mittelpunktschätzung

Wir gehen davon aus, daß die gemessenen Punkte annähernd über den gesamten Winkelbereich gleichverteilt sind. Andernfalls muß man modifizierte Verfahren verwenden.

5.1 Schwerpunktbildung

Die einfachste Methode zur a-priori-Mittelpunktschätzung ist sicherlich durch die Schwerpunktbildung gegeben. D.h.

$$\hat{x}_M = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{und} \quad \hat{y}_M = \frac{1}{n} \sum_{i=1}^{n} y_i .$$

Man kann dieses Verfahren auch als Regressionsverfahren interpretieren. Berechnet man nämlich für die Punkte $P_1,\ldots,P_n$ die zugehörige Regressionsgeraden in x-Richtung bzw. y-Richtung, d.h. man bestimmt zwei Geraden $g_1(t) = m_1t+n_1$ und $g_2(t) = m_2t+n_2$ durch

$$g_1 : \mathrm{Min}_{m_1,n_1} \sum_{i=1}^{n} (m_1x_i+n_1-y_i)^2$$

und

$$g_2 : \mathrm{Min}_{m_2,n_2} \sum_{i=1}^{n} (m_2x_i+n_2-y_i)^2 ,$$

so ist $(\hat{x}_M,\hat{y}_M)$ gerade der Schnittpunkt dieser beiden Geraden.

5.2 Min-Max-Bildung

Seien

$$x_{Min} = \mathrm{Min}\{x_i;\ i=1,\ldots,n\} \qquad x_{Max} = \mathrm{Max}\{x_i;\ i=1,\ldots,n\}$$

$$y_{Min} = \mathrm{Min}\{y_i;\ i=1,\ldots,n\} \qquad y_{Max} = \mathrm{Max}\{y_i;\ i=1,\ldots,n\} ,$$

dann ist eine mögliche a-priori-Mittelpunktschätzung gegeben durch

$$\hat{x}_M = x_{Min} + \frac{1}{2} (x_{Max} - x_{Min})$$

$$\hat{y}_M = y_{Max} + \frac{1}{2} (y_{Max} - y_{Min}) .$$

5.3 Modifizierte Min-Max-Bildung

Das einfache Min-Max-Verfahren läßt sich noch modifizieren. Dazu seien für $a_1,\ldots,a_n \in \mathbb{R}$ $a_{[k]}$, $k=1,\ldots,n$, die jeweils k-ten Elemente dieser Zahlen, der Größe nach, geordnet.

Beispiel:

Für $a_1 = 5,\ a_2 = 3,\ a_3 = 1,\ a_4 = 1,\ a_5 = 7$ gilt:

$$a_{[1]} = 1,\ a_{[2]} = 1,\ a_{[3]} = 3,\ a_{[4]} = 5,\ a_{[5]} = 7.$$

Für $k=1,\ldots,[\frac{n}{2}]$ bestimme man

$$x_M^{(k)} = x_{[k]} + \frac{1}{2}(x_{[n-k]} - x_{[k]})$$

und

$$y_M^{(k)} = y_{[k]} + \frac{1}{2}(y_{[n-k]} - y_{[k]}$$

und setze für eine Zahl $1 \quad N \quad [\frac{n}{2}]$

$$\hat{x}_M = \frac{1}{N} \sum_{k=1}^{N} x_M^{(k)} \quad \text{und} \quad \hat{y}_M = \frac{1}{N} \sum_{k=1}^{N} y_M^{(k)} .$$

6. Abschließende Bemerkung

Zum Abschluß möchte ich noch einmal auf die Behandlung des Problems mit dem zu Beginn vorgestellten Regressionsmodell zurückkommen. Die vorangegangenen Überlegungen haben gezeigt, daß man bei der Behandlung des in (1.1) gestellten Problems ohne Kenntnis der Winkel unter denen die Punkte gemessen wurden, einen gewissen Aufwand zur Lösung dieses Problems betreiben muß. Natürlich ist dieser Aufwand verständlich, da ohne Kenntnis der Winkel die Lösung von (1.1) zu einem nichtlinearen Problem führt, verbunden mit allen dabei bekanntlich auftretenden Schwierigkeiten. Insbesondere bedingt die Forderung, die gemessenen Punkte nicht unbedingt abspeichern zu wollen, gewisse Zusatzschwierigkeiten. All diese Schwierigkeiten konnten aber insgesamt - wie die vorangegangenen Ausführungen zeigen - sehr gut gelöst werden. Dennoch hat das angesprochene Regressionsverfahren den Vorteil, vollständig linear zu sein. Daher wäre darüber nachzudenken, ob nicht durch eine zusätzliche Meßeinrichtung oder durch gewisse Zusatzüberlegungen die Winkel der gemessenen Punkte mitbestimmt werden können. Dadurch würde sich die Lösung des Problems (1.1) erheblich vereinfachen.